

LoRA 기반 Fine-Tuning 을 통한 Fading 채널에서의 신속한 Semantics Alignment 알고리즘 연구

구준회, 김상혁, 고승우
인하대학교

jeremy0915@inha.edu, inhaaese20@inha.edu, swko@inha.ac.kr

Agile Semantics Alignment over Fading Channel via LoRA based Fine-Tuning

Joon-Hoe Koo, Sang-Hyeok Kim, Seung-Woo Ko
Inha University

요약

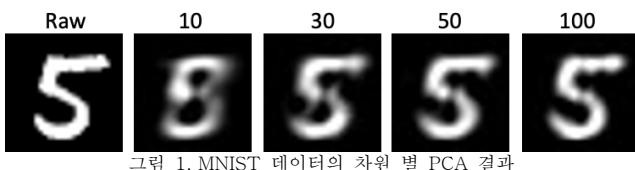
본 논문에서는 무선 채널 품질 악화에 따른 송수신기 간의 입력 Feature 벡터 차원 불일치로 인한 Semantics Misalignment를 빠르게 해결하기 위한 통신 및 학습 기법을 제안한다. 채널 변화에 따른 입력 데이터의 Feature 차원 및 학습 데이터 숫자를 제어하고, Low-Rank Adaptation (LoRA)를 사용한 Fine-Tuning 기법을 통해 Task를 신속히 해결한다. V2X 송수신기를 통한 실제 실험 결과, 제안 기법은 End-to-End Training 모델에 비해 비약적인 통신 및 시간 절감을 가능하게 하면서도 97.6%의 높은 분류 정확도를 보인다.

I. 서론

지능형 단말과 고밀도 사물이 연결되는 6G 시대에는 비트 단위의 전송보다는 의미 (Semantics) 기반 정보 전달의 효율이 통신의 품질을 결정한다 [1]. 그러나 무선 채널 품질은 수시로 변화하여 전송 가능한 Feature 벡터의 차원 수가 달라지고, 이로 인해 송신기 (TX)에서 보내는 차원과 수신기 (RX)의 Pre-Trained 모델 간의 Semantics Misalignment 가 발생하며 Task 성공률을 급격히 저하시킨다 [2]. 이러한 문제를 저지연으로 해결하기 위해, 본 논문에서는 기존 Pre-Trained 모델을 유지하면서도, 송신기에서의 Feature 벡터의 차원 제어 및 Fine-Tuning을 통해 Semantics Misalignment를 빠르게 해결한다.

II. 시스템 모델

본 논문에서는 Split 구조의 Semantic Communication을 고려한다. 이 때, TX 와 RX 하나의 쌍을 고려하는데, TX 측 인코더는 입력 이미지를 Principle Component Analysis (PCA)를 통해 N 차원 Feature 벡터로 압축하며, 채널 상태에 따라 보낼 수 있는 차원의 수를 제어할 수 있다. RX 측 디코더는 End-to-End Training 된 MLP 기반 모델을 보유하고, 수신한 PCA 벡터를 복원하여 분류하는데 사용한다. 채널 환경의 변화로 전송 차원이 달라지면 원본 데이터가 그림 1 과 같이 바뀌게 된다. 바뀐 이미지를 분류하기 위해선 새로운 모델의 재학습이 필요한데, 이 때 End-to-End Training은 많은 데이터와 시간을 필요로 하고 무선 채널 환경의 Variation에 효율적으로 대응이 불가능하다.



III. Low-Rank Adaptation (LoRA) 기반 Fine-Tuning

III. A를 통해 Dimension 변화에 따른 통신 Payload 구

성을 설명하고, III. B를 통해 LoRA 알고리즘에 대해 구체적으로 설명한다. III. C에서 TX, RX 송수신의 프레임워크에 대해 논의한다.

A. Dimension 변화에 따른 통신 Payload 구성

채널 상태 변화로 인해 TX에서 전송할 테스트 데이터셋의 차원이 감소하고, 차원이 감소함에 따라 통신 Payload에 여분의 용량이 생긴다. 여분의 용량에 Fine-Tuning 용 학습 데이터셋을 함께 전송하는데, 이 때 최대 전송용량을 B_{max} 로 정의할 때, 전송 가능한 학습 데이터셋의 최대 개수는 다음과 같다:

$$N_{max} = \left\lfloor \frac{B_{max} - H - t(h + \sigma\alpha)}{h + \sigma\alpha} \right\rfloor.$$

위 식에서 H 는 테스트 및 학습 데이터셋의 파일 전체의 헤더(8byte)이고, h 는 벡터 한개의 헤더(4byte)이다. t 는 테스트 데이터의 수, α 는 PCA를 진행할 차원의 수이며, σ 는 α 당 바이트 수(float 한개당 4byte)를 뜻한다. 이 식을 통해 TX는 PCA 차원마다 보낼 수 있는 학습 데이터셋의 수를 계산하여 테스트 데이터셋과 함께 RX로 송신한다.

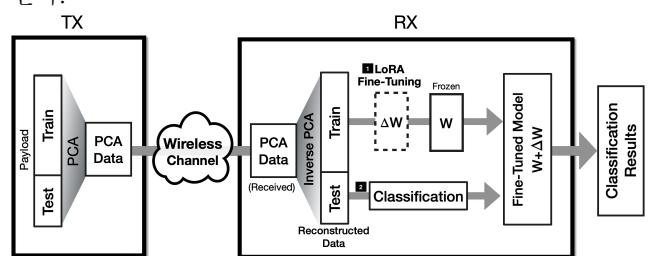


그림 2. 시스템 프레임워크 (RX에서 1 번 진행 후 2 번을 진행한다.)

B. Low-Rank Adaptation

LoRA[3]는 Pretrained Model의 원본 가중치 행렬 W 에 직접 손대지 않고, Low Rank 행렬 A, B 를 추가로 학습하는 Fine-Tuning 기법이다. W 대신 $\Delta W = AB$ 를 학습함으로써 학습 데이터셋과 Epoch를 줄이면서도 모델의 성능을 유지한다. 랭크 r 를 조절하여 ΔW 의 용량과 학습 복잡도를 제어하고, Pre-Trained 모델의 지식은 고정한 채로 새로운 차원의 특징에 빠르게 적응한다.

C. 데이터 송수신 프레임워크

그림 2는 본 시스템의 프레임워크를 나타낸다. 먼저 TX에서 분류할 테스트 데이터셋을 차원 α 로 PCA 하여 Feature 벡터를 추출한 후에, Payload 구성에 따라 학습 데이터셋의 최대 개수 N_{max} 를 계산하여 그 수 만큼 동일 차원으로 PCA를 진행한다. 그 후 각 데이터셋의 Feature 벡터들을 RX로 전송을 하면, RX 애선 테스트와 학습 데이터셋의 Feature 벡터들을 Inverse PCA 하여 이미지를 복원한다. 복원한 학습 데이터셋을 이용하여 Pre-Trained 모델을 LoRA Fine-Tuning으로 학습하고, 복원한 테스트 데이터셋을 학습시킨 모델에 입력하여 Classification을 진행한다.

IV. Experiments

TX, RX 역할을 하는 두 대의 SIRIUS [4] 기기와 각각 연결된 두대의 노트북을 약 2m 간격으로 배치한다. 기존 256QAM에서 64QAM의 변조방식으로 설정하여 채널 상태가 악화된 상황을 가정하였다. RX가 보유한 Pre-Trained 모델의 차원인 200 차원의 $3/4$ 인 150 차원을 시작으로 10 차원 단위로 실험을 진행하였다. RX는 수신한 학습 데이터셋을 이용하여 Pre-Trained 모델을 내장 CPU를 통해 Fine-Tuning하여 학습시킨 후, 수신된 테스트 데이터셋을 Fine-Tuned 모델에 입력하여 실험의 Task인 MNIST 데이터셋의 분류 성능을 검증한다.

Latency 측정의 경우, TX, RX 역할을 하는 각 SIRIUS 기기가 GPS 수신기로부터 Pulse Per Second (PPS) 신호로 동기화된 시간을 출력한다. TX에서 PCA를 시작하는 순간의 시각을 A, RX에서 데이터를 수신한 뒤에 Image Reconstruction이 완료된 순간의 시각을 B, Fine-Tuning으로 Pre-Trained 모델을 학습하는데 걸리는 시간을 C, 분류에 걸리는 시간을 D라고 한다면, 전체 Latency를 나타내는 식은 다음과 같이 정의할 수 있다.

$$\text{End-to-End Latency} = (B - A) + C + D.$$

위 식에서 Computing Latency에 해당하는 C와 D를 제외한 $(B - A)$ 를 PCA에 걸리는 시간이 짧다는 점을 고려하여 Communication Latency로 간주할 수 있다.

그림 3은 각 차원에서의 세 가지 Latency들을 출력한 결과이다. 낮은 차원으로 PCA를 진행할수록 테스트 데이터셋의 용량이 적어져 더 많은 학습 데이터셋을 RX로 전송할 수 있는데, 그만큼 Fine-Tuning을 진행할 때의 데이터 양이 많아져 더 오랜 시간이 걸리기 때문에 α 가 높아질수록 Fine-Tuning에 걸리는 시간이 줄어드는 경향성을 보인다. 그림 4는 각 차원을 사용했을 때 End-to-End Training 모델과 LoRA Fine-Tuning 된 Pre-Trained 모델의 분류 정확도를 비교한다. $\alpha = 40$ 일 때를 제외한 모든 차원에서 LoRA Fine-Tuning을 한 모델이 더 우수한 성능을 보이는 것을 확인하였다.

V. 결론

본 논문에서 무선 채널의 품질이 악화되어 전송 가능한 Feature 벡터의 수가 줄었을 때 TX와 RX 간에 발생한 Semantics Misalignment를 신속히 해결할 수 있는 통신 및 학습 기법을 제안한다. 기존의 Backbone Model을 LoRA Fine-Tuning을 통해 기존 End-to-End Training 모델의 성능을 유지하였다. 추후에는 제안하는 알고리즘을 다양한 데이터셋과 채널 환경에서 검증하고 최적의 차원 제어 기법에 대해 연구할 예정이다.

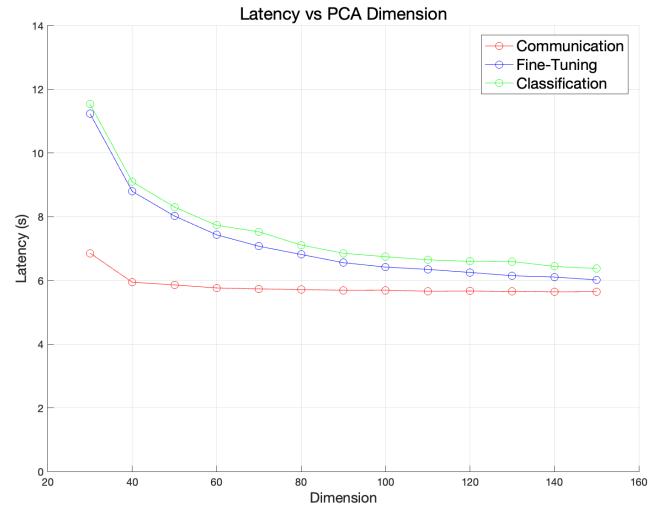


그림 3. Latency-Dimension 그래프

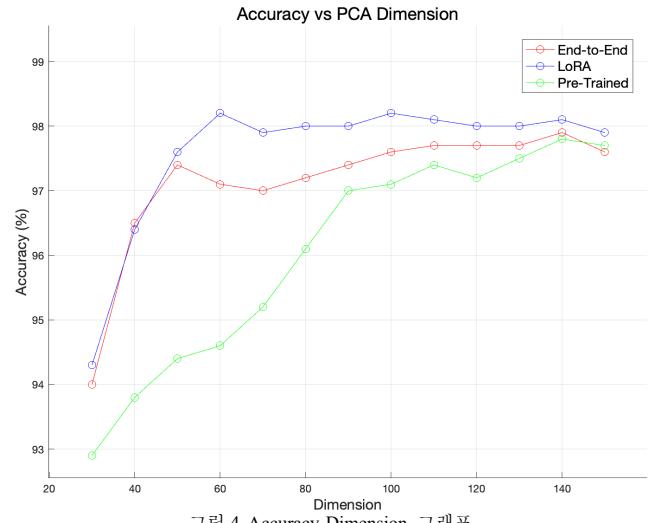


그림 4. Accuracy-Dimension 그래프

ACKNOWLEDGMENT

이 논문은 2025년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00453301).

참고 문헌

- [1] Qin, Z., Tao, X., Lu, J., Tong, W., & Li, G. Y. (2021). Semantic communications: Principles and challenges. arXiv preprint arXiv:2201.01389.
- [2] J. Choi, J. Park, S. -W. Ko, J. Choi, M. Bennis and S. -L. Kim, "Semantics Alignment via Split Learning for Resilient Multi-User Semantic Communication," in IEEE Transactions on Vehicular Technology, vol. 73, no. 10, pp. 15815-15819, Oct. 2024, doi: 10.1109/TVT.2024.3410380.
- [3] A. Tiwary et al., "Fine-Tuning Vision Transformer Using LoRA for Image Classification," in Proc. 8th Int. Conf. Electron. Mater. Eng. Nano-Tech (IEMENTech), 2025, pp. 1-4, doi:10.1109/IEMENTech65115.2025.10959407.
- [4] Ettifos, "SIRIUS SDR Platform," Ettifos, 2025. [Online]. Available: <https://www.ettifos.com/product-sirius>. [Accessed: May 16, 2025].